

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 16:12:33 ; Search time 42 Seconds
(without alignments)

200.298 Million cell updates/sec

Title: US-09-991-003B-8

Perfect score: 276

Sequence: 1 DNDGDGDDNDGGGNGDGG.....MMVVRIMVVMVIMLLMT 53

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	276	100.0	53	ABB79541	Rhinovirus pertub
2	135	48.9	78	ABG48827	Human liver peptid
3	135	48.9	78	ABB28821	Peptide #1472 enco
4	135	48.9	78	ABB34006	Peptide #1512 enco
5	135	48.9	78	ABB19447	Protein #1446 enco
6	135	48.9	78	AAW54774	Human brain expres
7	135	48.9	78	AAW67161	Human bone marrow
8	135	48.9	78	AAW15020	Peptide #1454 enco
9	135	48.9	78	AAW27458	Peptide #1495 enco

10	135	48.9	78	22	AAW02750	Peptide #1432 enco
11	135	48.9	78	23	ABG36823	Human peptid enco
12	126	45.7	402	22	ABG12558	Novel human diagno
13	115.5	41.8	349	22	ABG00656	Novel human diagno
14	114	41.3	426	22	ABG19098	Novel human diagno
15	109	39.5	105	22	ABG48281	Human liver peptid
16	109	39.5	105	22	ABG28259	Human peptid #910
17	109	39.5	105	22	ABG33434	Peptide #940 enco
18	109	39.5	105	22	ABG18893	Protein #892 enco
19	109	39.5	105	22	AAW54219	Human brain expres
20	109	39.5	105	22	AAW66613	Human bone marrow
21	109	39.5	105	22	AAW14486	Peptide #920 enco
22	109	39.5	105	22	AAW26899	Peptide #936 enco
23	109	39.5	105	22	AAW02213	Peptide #895 enco
24	109	39.5	105	23	ABG36265	Human peptid enco
25	101	36.6	252	22	ABG04689	Novel human diagno
26	97.5	35.3	2747	22	ABG02342	Novel human diagno
27	94	34.1	711	22	ABG02174	Novel human diagno
28	93.5	33.9	53	22	ABG48282	Human liver peptid
29	93.5	33.9	53	22	ABG28260	Human peptid #911
30	93.5	33.9	53	22	ABG33435	Peptide #941 enco
31	93.5	33.9	53	22	ABG18894	Protein #893 enco
32	93.5	33.9	53	22	AAW54220	Human brain expres
33	93.5	33.9	53	22	AAW66614	Human bone marrow
34	93.5	33.9	53	22	AAW14487	Peptide #921 enco
35	93.5	33.9	53	22	AAW26900	Peptide #937 enco
36	93.5	33.9	53	22	AAW02214	Peptide #896 enco
37	93.5	33.9	53	23	ABG36266	Human peptid enco
38	91	33.0	138	22	ABG50800	Human liver peptid
39	91	33.0	138	22	ABG30765	Peptide #3416 enco
40	91	33.0	138	22	ABG35941	Peptide #3447 enco
41	91	33.0	138	22	ABG21351	Protein #3350 enco
42	91	33.0	138	22	AAW56743	Human brain expres
43	91	33.0	138	22	AAW69125	Human bone marrow
44	91	33.0	138	22	AAW16955	Peptide #3389 enco
45	91	33.0	138	22	AAW29444	Peptide #3381 enco

ALIGNMENTS

RESULT 1

ABB79541
ID ABB79541 standard; Protein; 53 AA.

AC ABB79541;

DT 01-OCT-2002 (first entry)

XX Rhinovirus pertubagen cw985.

XX cw985; pertubagen; rhinovirus; viral infection; virucide; human;

XX gene therapy.

XX Homo sapiens.

XX WO200255697-A2.

XX 18-JUL-2002.

XX 16-NOV-2001; 2001WO-US43486.

XX 27-NOV-2000; 2000US-253333P.

XX 28-FEB-2001; 2001US-272026P.

XX (DELT-) DELTAGEN PROTEOMICS INC.

XX Kamb CA, Poritz MA, Teng DH;

XX WPI; 2002-557822/59.

XX N-PSDB; ABN84303.

XX New cw985 pertubagen polypeptides and polynucleotides useful for

PT treating viral infections by the picornaviridae class, in chromosomal
 PT mapping, tissue typing, forensic biology, or viral serotyping -
 PS
 PS Claim 1; Fig 11; 117pp; English.

XX The present sequence is the protein sequence of novel cW985
 CC pertubagen. To identify molecules that alter a cell's ability to
 CC resist or deter rhinovirus (RV) infection, a random primed library
 CC of 12 x 10 power 6 clones was constructed from cDNA isolated from
 CC placental tissue. This genetic library was transfected into HeLa
 CC cells that were previously shown to be susceptible to RV infection.
 CC The library-containing population was expanded 9-fold and then
 CC exposed to human RV-14 (ATCC VR 284) to identify perturbagens that
 CC inhibited the viral life cycle. Clone cW985 was selected. The
 CC invention further provides host cells, vectors and gene therapy
 CC vectors comprising polynucleotides encoding cW985. The host cells
 CC provide for methods for producing polypeptides having viral-related
 CC activity, which in turn can be used to identify potential
 CC therapeutics. The invention also provides methods for identifying
 CC a cellular target that interacts with the pertubagen, e.g. using a
 CC yeast two-hybrid assay, and for using such targets to screen for
 CC putative viral therapeutics. Pertubagen polynucleotides,
 CC polypeptides or antibodies may also be used in detection assays
 CC (e.g. chromosomal mapping, tissue typing, forensic biology, or
 CC viral serotyping), and in predictive medicine.

XX Sequence 53 AA;
 SQ
 Query Match 100.0%; Score 276; DB 23; Length 53;
 Best Local Similarity 100.0%; Pred. No. 1.4e-26;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DNDGDDNDGGGNGGSDNYATDDMMVMVMVMVMVMVMVMVMIIMLLMT 53
 Db 1 DNDGDDNDGGGNGGSDNYATDDMMVMVMVMVMVMVMVMVMIIMLLMT 53

RESULT 2

ABG48827
 ID ABG48827 standard; Peptide; 78 AA.

XX
 AC ABG48827;

XX 25-FEB-2003 (first entry)

XX Human liver peptide, SEQ ID No 27475.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
 KW hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00664.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488898/53.

PT Human genome-derived single exon nucleic acid probes useful for

PT analysing gene expression in human adult liver -
 XX Claim 27; SEQ ID No 27475; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
 CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 1109 defined nucleotide sequences given in the
 CC specification (or complements/ fragments). The probe hybridises at high
 CC stringency to a nucleic acid molecule expressed in the human adult
 CC liver. (I) may be used for predicting, measuring and displaying gene
 CC expression in samples derived from human adult liver. The genes
 CC identified may be involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
 CC is associated with coronary heart disease. ABG47348-ABG59930 represent
 CC human liver single exon encoded peptides of the invention.

CC Note: The sequence information for this patent does not appear in the
 CC printed specification but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 78 AA;

Query Match 48.9%; Score 135; DB 22; Length 78;

Best Local Similarity 53.6%; Pred. No. 3.2e-09;

Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;

OY 1 DNDGDDNDGGGNGGSDNYATDD--MMVMVMVM--MVVVRIMVMVMIIMLLM 52

Db 12 DDDGDDDDDDADNDGDDGDDGDDDDVTMMVMVMVMVMVMVMIIMVMM 67

RESULT 3

ABB28821

ID ABB28821 standard; Peptide; 78 AA.

XX
 AC ABB28821;

XX 01-FEB-2002 (first entry)

XX Peptide #1472 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID No 11789; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived

CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 78 AA;
Query Match 48.9%; Score 135; DB 22; Length 78;
Best Local Similarity 53.6%; Pred. No. 3.2e-09;
Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;

QY 1 DNDGGDNDGGGNGNG-GGDNYATDD--MMVMIMM-MVVVRIMVVMVMIILLM 52
|:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:|
Db 12 DDDGGDDSDDDADNDGGDGGDGGDDVTMMVMIMMAMVMVMVMIIVMM 67

RESULT 4
ABB34006
ID ABB34006 standard; Peptide; 78 AA.

XX AC ABB34006;
XX DT 04-FEB-2002 (first entry)

DE Peptide #1512 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.

XX WO200157277-A2.
XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human foetal liver -

XX Claim 27; SEQ ID NO 26641; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 78 AA;

Query Match 48.9%; Score 135; DB 22; Length 78;
Best Local Similarity 53.6%; Pred. No. 3.2e-09;
Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;

QY 1 DNDGGDNDGGGNGNG-GGDNYATDD--MMVMIMM-MVVVRIMVVMVMIILLM 52
|:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:|
Db 12 DDDGGDDSDDDADNDGGDGGDGGDDVTMMVMIMMAMVMVMVMIIVMM 67

RESULT 5
ABB19447
ID ABB19447 standard; Protein; 78 AA.

XX AC ABB19447;
XX DT 23-JAN-2002 (first entry)

XX Protein #1446 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.

XX OS Homo sapiens.
XX PN WO200157274-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -

XX Claim 15; SEQ ID No 21217; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 78 AA;

Query Match 48.9%; Score 135; DB 22; Length 78;

```
Best Local Similarity 53.6%; Pred. No. 3.2e-09;
Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;

OY 1 DNDGDDNDGGGNGG-GGDNATDD--MMVMIMM-MVVVRIMVVMVIMLILM 52
Db 12 DDDGDDSDDDADNDGDDGDDGDDVTMMVMIMMAMVMVMVMVIMVMM 67

RESULT 6
AAM54774
ID AAM54774 standard; Protein; 78 AA.
XX
AC AAM54774;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26879.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-483446/52.
XX
Single exon nucleic acid probes for analyzing gene expression in human
brains -
XX
Example 4; SEQ ID NO: 26879; 650pp + Sequence Listing; English.
XX
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human
brain. They can be used to measure gene expression in brain cell samples,
which may enable the diagnosis and improved treatment of nervous system
diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
epilepsy and cancers. The present sequence is a protein encoded by one of
the probes of the invention.
XX
Sequence 78 AA;
XX
Query Match 48.9%; Score 135; DB 22; Length 78;
Best Local Similarity 53.6%; Pred. No. 3.2e-09;
Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;
OY 1 DNDGDDNDGGGNGG-GGDNATDD--MMVMIMM-MVVVRIMVVMVIMLILM 52
Db 12 DDDGDDSDDDADNDGDDGDDGDDVTMMVMIMMAMVMVMVIMVMM 67

RESULT 7
AAM67161
ID AAM67161 standard; Protein; 78 AA.
XX
AC AAM67161;
XX
```

```
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27467.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488900/53.
XX
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human bone marrow -
XX
Example 4; SEQ ID NO: 27467; 658pp + Sequence Listing; English.
XX
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human
bone marrow. They can be used to measure gene expression in bone marrow
samples, which may enable the improved diagnosis and treatment of cancers
such as lymphoma, leukaemia and myeloma. The present sequence is a
protein encoded by one of the probes of the invention.
XX
Sequence 78 AA;
XX
Query Match 48.9%; Score 135; DB 22; Length 78;
Best Local Similarity 53.6%; Pred. No. 3.2e-09;
Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;
OY 1 DNDGDDNDGGGNGG-GGDNATDD--MMVMIMM-MVVVRIMVVMVIMLILM 52
Db 12 DDDGDDSDDDADNDGDDGDDGDDVTMMVMIMMAMVMVMVIMVMM 67

RESULT 8
AAM15020
ID AAM15020 standard; Protein; 78 AA.
XX
AC AAM15020;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #1454 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
```

```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 27; SEQ ID No 19846; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs; see AA110068-AA128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence . 78 AA;

Query Match 48.9%; Score 135; DB 22; Length 78;
Best Local Similarity 53.6%; Pred. No. 3.2e-09;
Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;

Oy 1 DNDGDDNDGGGNGG-GGDNATDD--MMVMIMM-MVVRIMVMVMIIMLLM 52
|:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:|
Db 12 DDDGDDSDSDADNDGDDGDDGDDVTMTVMIMMAMVMVMVMIIMVMM 67

RESULT 9
AAW27458
ID AAW27458 standard; Protein; 78 AA.
XX
XX AAW27458;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #1495 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 27; SEQ ID No 11490; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX (see AA100010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for

```

```

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488997/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID No 27727; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs;
XX see AA131315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX
XX Sequence 78 AA;

Query Match 48.9%; Score 135; DB 22; Length 78;
Best Local Similarity 53.6%; Pred. No. 3.2e-09;
Matches 30; Conservative 13; Mismatches 9; Indels 4; Gaps 3;

Oy 1 DNDGDDNDGGGNGG-GGDNATDD--MMVMIMM-MVVRIMVMVMIIMLLM 52
|:|||||:| |:|||||:| |:|||||:| |:|||||:| |:|||||:|
Db 12 DDDGDDSDSDADNDGDDGDDGDDVTMTVMIMMAMVMVMVMIIMVMM 67

RESULT 10
AAW02750
ID AAW02750 standard; Protein; 78 AA.
XX
XX AAW02750;
XX
XX 09-OCT-2001 (first entry)
XX
XX Peptide #1432 encoded by probe for measuring breast gene expression.
XX
XX Probe; human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 27; SEQ ID No 11490; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX (see AA100010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for

```


APPLICATION NUMBER: US/08/483,432
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/361,920
FILING DATE:
APPLICATION NUMBER: US 07/940,860
FILING DATE: 28-OCT-1992
APPLICATION NUMBER: DK 1158/90
FILING DATE: 09-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00124
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3435.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Humicola insolens
STRAIN: DSM 1800
US-08-483-432-14

Query Match 25.0%; Score 69; DB 1; Length 42;
Best Local Similarity 52.4%; Pred. No. 0.06;
Matches 11; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 2 NDDGDDNDGGGNNGGGDN 22
Db 3 SNNGGNNGGGNNGGGNN 23

RESULT 5
PCT-US92-06840-2
Sequence 2, Application PC/TUS9206840
GENERAL INFORMATION:
APPLICANT: Shi, Yang
APPLICANT: Seto, Edward
APPLICANT: Shenk, Thomas
TITLE OF INVENTION: Y11 TRANSCRIPTION FACTOR AND METHODS OF
ISOLATING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas - 7th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06840
FILING DATE: 19920814
CLASSIFICATION:
AU 1805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/746,485
FILING DATE: 16-AUG-1991
ATTORNEY/AGENT INFORMATION:

NAME: Dennis, Manette
REGISTRATION NUMBER: 30,623
REFERENCE/DOCKET NUMBER: M-12594 CIP (1570-8)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-06840-2

Query Match 25.0%; Score 69; DB 5; Length 414;
Best Local Similarity 45.8%; Pred. No. 0.88;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 1 DNDGDDNDGGGNNGGGDN 24
Db 45 EEDDDDEDGGGDDGGGGHGHGA 68

RESULT 6
US-08-681-129-2
Sequence 2, Application US/08681129
Patent No. 5738854
GENERAL INFORMATION:
APPLICANT: Mettenleiter, Thomas Cristoph
TITLE OF INVENTION: Pseudorabies virus vaccine
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330 Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,129
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/244,446
FILING DATE: 02-JUN-1994
CLASSIFICATION: 424
APPLICATION NUMBER: EP 92.203.079.6
FILING DATE: 06-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
STRAIN: Kaplan
US-08-681-129-2

Query Match 24.1%; Score 66.5; DB 1; Length 525;
Best Local Similarity 28.3%; Pred. No. 2.3;

Length 764;
; Indels

SCHOOL
OF
AY

30

```
;
;
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 764 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-16930-4

Query Match      23.9%; Score 66; DB 5; Length 764;
Best Local Similarity 48.3%; Pred. No. 4.1;
Matches 14; Conservative 5; Mismatches 8; Indels 2; Gaps 1;

Oy 1 DNDGDDNDG--GGENNGGSDNYATDD 27
Db 574 DDDDDDDDDGEGDEDDDDDDDDDD 602

RESULT 10
US-08-375-300-2
; Sequence 2, Application US/08375300
; Patent No. 5679566
; GENERAL INFORMATION:
; APPLICANT: Feng, He
; APPLICANT: Jacobson, Allan S.
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
; THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,300
; FILING DATE: 20-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. P.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04020/046001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)542-5070
; TELEFAX: (617)542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-375-300-2

Query Match      23.9%; Score 66; DB 1; Length 1089;
Best Local Similarity 48.3%; Pred. No. 6.3;
Matches 14; Conservative 5; Mismatches 8; Indels 2; Gaps 1;

Oy 1 DNDGDDNDG--GGENNGGSDNYATDD 27
Db 899 DDDDDDDDDGEGDEDDDDDDDDDD 927

RESULT 11
US-09-177-431-2
; Sequence 2, Application US/09177431
; Patent No. 6071700
; GENERAL INFORMATION:
; APPLICANT: He, Feng
; APPLICANT: Jacobson, Allan S.
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
; ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/177,431
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/955,472
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/050001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-9806
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-177-431-2

Query Match      23.9%; Score 66; DB 3; Length 1089;
Best Local Similarity 48.3%; Pred. No. 6.3;
Matches 14; Conservative 5; Mismatches 8; Indels 2; Gaps 1;

Oy 1 DNDGDDNDG--GGENNGGSDNYATDD 27
Db 899 DDDDDDDDDGEGDEDDDDDDDDDD 927

RESULT 12
PCT-US95-16930-2
; Sequence 2, Application PC/TUS9516930
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE
; PRODUCTION IN THE ABSENCE OF
; NONSENSE-MEDIATED mRNA DECAY
; FUNCTION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

1	APPLICATION NUMBER: PCT/US95/16930	
1	FILING DATE: 27-DEC-1995	
1	PRIOR APPLICATION DATA:	
1	APPLICATION NUMBER: US 08/375,300	
1	FILING DATE: 20-JAN-1995	
1	ATTORNEY/AGENT INFORMATION:	
1	NAME: Fasse, J. Peter	
1	REGISTRATION NUMBER: 32,983	
1	REFERENCE/DOCKET NUMBER: 04020/046W01	
1	TELECOMMUNICATION INFORMATION:	
1	TELEPHONE: (617)542-5070	
1	TELEFAX: (617)542-8906	
1	TELEX: 200154	
1	INFORMATION FOR SEQ ID NO: 2:	
1	SEQUENCE CHARACTERISTICS:	
1	LENGTH: 1089 amino acids	
1	TYPE: amino acid	
1	STRANDEDNESS: not relevant	
1	TOPOLOGY: linear	
1	MOLECULE TYPE: protein	
1	PCT-US95-16930-2	
1	Query Match 23.9%; Score 66; DB 5; Length 1089;	
1	Best Local Similarity 48.3%; Pred. No. 6.3;	
1	Matches 14; Conservative 5; Mismatches 8; Indels 2; Gaps 1;	
1	QY 1 DNDGDDNDG--GGENNGGDDNYATDD 27	
1	DB 899 DDDDDDDGEGDEDDDEDDDD 927	
1	RESULT 13	
1	US-08-560-398-8	
1	Sequence 8, Application US/08560398	
1	Patent No. 5907082	
1	GENERAL INFORMATION:	
1	APPLICANT: O'Neill, Sharman	
1	APPLICANT: Nadeau, Jeanette	
1	TITLE OF INVENTION: Ovule-Specific Gene Expression	
1	NUMBER OF SEQUENCES: 13	
1	CORRESPONDENCE ADDRESS:	
1	ADDRESSEE: Townsend and Townsend and Crew LLP	
1	STREET: Two Embarcadero Center, Eighth Floor	
1	CITY: San Francisco	
1	STATE: California	
1	COUNTRY: USA	
1	ZIP: 94111-3834	
1	COMPUTER READABLE FORM:	
1	MEDIUM TYPE: Floppy disk	
1	COMPUTER: IBM PC compatible	
1	OPERATING SYSTEM: PC-DOS/MS-DOS	
1	SOFTWARE: PatentIn Release #1.0, Version #1.30	
1	CURRENT APPLICATION DATA:	
1	APPLICATION NUMBER: US/08/560,398	
1	FILING DATE: 17-NOV-1995	
1	CLASSIFICATION: 435	
1	ATTORNEY/AGENT INFORMATION:	
1	NAME: Bastian, Kevin L.	
1	REGISTRATION NUMBER: 34,774	
1	REFERENCE/DOCKET NUMBER: 023070-063300US	
1	TELECOMMUNICATION INFORMATION:	
1	TELEPHONE: (415) 576-0200	
1	TELEFAX: (415) 576-0300	
1	INFORMATION FOR SEQ ID NO: 8:	
1	SEQUENCE CHARACTERISTICS:	
1	LENGTH: 190 amino acids	
1	TYPE: amino acid	
1	TOPOLOGY: linear	
1	MOLECULE TYPE: protein	
1	US-08-560-398-8	
1	Query Match 23.7%; Score 65.5; DB 2; Length 190;	
1	Best Local Similarity 46.4%; Pred. No. 0.93;	
1	Matches 14; Conservative 5; Mismatches 8; Indels 2; Gaps 1;	
1	QY 1 DNDGDDNDG--GGENNGGDDNYATDD 27	
1	DB 899 DDDDDDDGEGDEDDDEDDDD 927	
1	RESULT 13	
1	US-08-560-398-8	
1	Sequence 8, Application US/08560398	
1	Patent No. 5907082	
1	GENERAL INFORMATION:	
1	APPLICANT: O'Neill, Sharman	
1	APPLICANT: Nadeau, Jeanette	
1	TITLE OF INVENTION: Ovule-Specific Gene Expression	
1	NUMBER OF SEQUENCES: 13	
1	CORRESPONDENCE ADDRESS:	
1	ADDRESSEE: Townsend and Townsend and Crew LLP	
1	STREET: Two Embarcadero Center, Eighth Floor	
1	CITY: San Francisco	
1	STATE: California	
1	COUNTRY: USA	
1	ZIP: 94111-3834	
1	COMPUTER READABLE FORM:	
1	MEDIUM TYPE: Floppy disk	
1	COMPUTER: IBM PC compatible	
1	OPERATING SYSTEM: PC-DOS/MS-DOS	
1	SOFTWARE: PatentIn Release #1.0, Version #1.30	
1	CURRENT APPLICATION DATA:	
1	APPLICATION NUMBER: US/08/560,398	
1	FILING DATE: 17-NOV-1995	
1	CLASSIFICATION: 435	
1	ATTORNEY/AGENT INFORMATION:	
1	NAME: Bastian, Kevin L.	
1	REGISTRATION NUMBER: 34,774	
1	REFERENCE/DOCKET NUMBER: 023070-063300US	
1	TELECOMMUNICATION INFORMATION:	
1	TELEPHONE: (415) 576-0200	
1	TELEFAX: (415) 576-0300	
1	INFORMATION FOR SEQ ID NO: 8:	
1	SEQUENCE CHARACTERISTICS:	
1	LENGTH: 190 amino acids	
1	TYPE: amino acid	
1	TOPOLOGY: linear	
1	MOLECULE TYPE: protein	
1	US-08-560-398-8	
1	Query Match 23.7%; Score 65.5; DB 2; Length 190;	
1	Best Local Similarity 46.4%; Pred. No. 0.93;	
1	Matches 14; Conservative 5; Mismatches 8; Indels 2; Gaps 1;	
1	QY 1 DNDGDDNDG--GGENNGGDDNYATDD 27	
1	DB 899 DDDDDDDGEGDEDDDEDDDD 927	
1	RESULT 13	
1	US-08-560-398-8	
1	Sequence 8, Application US/08560398	
1	Patent No. 5907082	
1	GENERAL INFORMATION:	
1	APPLICANT: O'Neill, Sharman	
1	APPLICANT: Nadeau, Jeanette	
1	TITLE OF INVENTION: Ovule-Specific Gene Expression	
1	NUMBER OF SEQUENCES: 13	
1	CORRESPONDENCE ADDRESS:	
1	ADDRESSEE: Townsend and Townsend and Crew LLP	
1	STREET: Two Embarcadero Center, Eighth Floor	
1	CITY: San Francisco	
1	STATE: California	
1	COUNTRY: USA	
1	ZIP: 94111-3834	
1	COMPUTER READABLE FORM:	
1	MEDIUM TYPE: Floppy disk	
1	COMPUTER: IBM PC compatible	
1	OPERATING SYSTEM: PC-DOS/MS-DOS	
1	SOFTWARE: PatentIn Release #1.0, Version #1.30	
1	CURRENT APPLICATION DATA:	
1	APPLICATION NUMBER: US/08/560,398	
1	FILING DATE: 17-NOV-1995	

```

1  ADDRESSEE: Townsend and Townsend and Crew LLP
2  STREET: Two Embarcadero Center, Eighth Floor
3  CITY: San Francisco
4  STATE: California
5  COUNTRY: USA
6  ZIP: 94111-3834
7
8  COMPUTER READABLE FORM:
9
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patent In Release #1.0, Version #1.30
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/526/813
16 FILING DATE: 11-SEP-1995
17 CLASSIFICATION: 435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US 08/258,188
20 FILING DATE: 14-JUN-1994
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Garrett-Wackowski, Eugenia
23 REGISTRATION NUMBER: 37,330
24 REFERENCE/DOCKET NUMBER: 15280-206-1
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: (415) 576-0200
27 TELEFAX: (415) 576-0300
28 INFORMATION FOR SEQ ID NO. 2:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 758 amino acids
31 TYPE: amino acid
32 TOPOLOGY: linear
33 MOLECULE TYPE: protein
34
35 US-08-526-813-2

```

Query Match	Score 65;	DB 1;	Length 758;
23.6%			

Qy 3 DDGD-----NDD---GGGEN---NGGDGDNYAT 25
||| ||| ||| ||| : ||| : |||
Db 505 DDGDDYISGGQDDQLFGSGGNDKLSGGDGNVLT 539

Search completed: December 3, 2003, 16:20:39
Job time : 22 secs

;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 34191
;; LENGTH: 105
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL078584.17
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.4
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 11
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.5
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 8.8
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.6
US-09-864-761-34191

Query Match 39.5%; Score 109; DB 9; Length 105;
Best Local Similarity 35.2%; Pred. No. 4.5e-05;
Matches 25; Conservative 16; Mismatches 10; Indels 20; Gaps 3;
QY 2 NDGDDND-----GGGNNNG-----DGDNYATDDMMVMIMMMVVV-----RI 41
DB 5 DDDYDDTDDGGGGGSGTGGSDSGSDYRDNNCGGKMMIIMMIVVVVVVLMV 64
QY 42 MVMVMVIMLLM 52
DB 65 MVMVMVVVMV 75

RESULT 5
US-10-017-161-2048
; Sequence 2048, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2048
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-2048
Query Match 39.3%; Score 108.5; DB 12; Length 276;
Best Local Similarity 38.1%; Pred. No. 0.00014;
Matches 24; Conservative 10; Mismatches 18; Indels 11; Gaps 1;

QY 1 DNDGDDNDGGGNNNGGDNATDDMMVMIM-----MMVVVRIMVVVMVIM 49
DB 38 DGGGGSGGGGGGNGRSGDGRVAVMVMVMVMEVMVMAMLVEMVVVVVVVVVVVV 97
QY 50 LLM 52
DB 98 LVV 100

RESULT 6
US-10-017-161-1820
; Sequence 1820, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1820
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1820

Query Match 36.2%; Score 100; DB 12; Length 410;
Best Local Similarity 27.8%; Pred. No. 0.0018;
Matches 25; Conservative 13; Mismatches 12; Indels 40; Gaps 2;
QY 3 DGDGDDNDGGGNNNG-----GGGNNNG-----DGDNYATDDMMVMV 32
DB 115 DGGGGDCCGNDGCGGCGSVHDGSGDGGHVGVGIGVGDGCGGNHGMIV 174
QY 33 IM-----MMVVVRIMVVVMVIMLLM 52
DB 175 MVMVMVMTVEGMVMVMVIMVIMVIMVLV 204

RESULT 7
US-10-017-161-1782
; Sequence 1782, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1782
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1782

Query Match 35.0%; Score 96.5; DB 12; Length 329;
Best Local Similarity 38.2%; Pred. No. 0.0036;
Matches 21; Conservative 14; Mismatches 9; Indels 11; Gaps 2;


```

; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42285
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023000.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
US-09-864-761-42285

Query Match      29.0%; Score 80; DB 9; Length 69;
Best Local Similarity 56.0%; Pred. No. 0.047;
Matches 14; Conservative 10; Mismatches 1; Indels 1; Gaps 0;

Qy   28 MMVMVMMVVVRIMVVMVVIIMLMT 52
    |||||::|||::|||::|||::|||::|||::|||
Db   25 MMVMVMMVVIVMVMVMMVMMVMMVMM 49
    |||||::|||::|||::|||::|||::|||::|||

RESULT 13
US-10-017-161-2180
; Sequence 2180, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152

CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42285
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023000.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
US-09-864-761-42285

Query Match      29.0%; Score 80; DB 9; Length 69;
Best Local Similarity 56.0%; Pred. No. 0.047;
Matches 14; Conservative 10; Mismatches 1; Indels 1; Gaps 0;

Qy   28 MMVMVMMVVVRIMVVMVVIIMLMT 52
    |||||::|||::|||::|||::|||::|||::|||
Db   25 MMVMVMMVVIVMVMVMMVMMVMMVMM 49
    |||||::|||::|||::|||::|||::|||::|||

RESULT 14
US-10-029-386-27424
; Sequence 27424, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27424
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR22_229.0
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
US-10-029-386-27424

Query Match      28.1%; Score 77.5; DB 12; Length 59;
Best Local Similarity 47.1%; Pred. No. 0.075;
Matches 16; Conservative 4; Mismatches 7; Indels 7; Gaps 1;

Qy   1 DNDGDDNDDG-----GGENGGDGDNYATDD 27
    |||||||::|||::|||::|||::|||::|||
Db   19 DNDGDDGDDGDDVDGDDGDDGDDGDDGDDGDD 52
    |||||||::|||::|||::|||::|||::|||

RESULT 15
US-09-864-761-35629
; Sequence 35629, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23

```

1	PRIOR APPLICATION NUMBER: US 60/180,312	
2	PRIOR FILING DATE: 2000-02-04	
3	PRIOR APPLICATION NUMBER: US 60/207,456	
4	PRIOR FILING DATE: 2000-05-26	
5	PRIOR APPLICATION NUMBER: US 09/632,366	
6	PRIOR FILING DATE: 2000-08-03	
7	PRIOR APPLICATION NUMBER: GB 24263.6	
8	PRIOR FILING DATE: 2000-10-04	
9	PRIOR APPLICATION NUMBER: US 60/236,359	
10	PRIOR FILING DATE: 2000-09-27	
11	PRIOR APPLICATION NUMBER: PCT/US01/00666	
12	PRIOR FILING DATE: 2001-01-30	
13	PRIOR APPLICATION NUMBER: PCT/US01/00667	
14	PRIOR FILING DATE: 2001-01-30	
15	PRIOR APPLICATION NUMBER: PCT/US01/00664	
16	PRIOR FILING DATE: 2001-01-30	
17	PRIOR APPLICATION NUMBER: PCT/US01/00669	
18	PRIOR FILING DATE: 2001-01-30	
19	PRIOR APPLICATION NUMBER: PCT/US01/00665	
20	PRIOR FILING DATE: 2001-01-30	
21	PRIOR APPLICATION NUMBER: PCT/US01/00668	
22	PRIOR FILING DATE: 2001-01-30	
23	PRIOR APPLICATION NUMBER: PCT/US01/00663	
24	PRIOR FILING DATE: 2001-01-30	
25	PRIOR APPLICATION NUMBER: PCT/US01/00662	
26	PRIOR FILING DATE: 2001-01-30	
27	PRIOR APPLICATION NUMBER: PCT/US01/00661	
28	PRIOR FILING DATE: 2001-01-30	
29	PRIOR APPLICATION NUMBER: PCT/US01/00670	
30	PRIOR FILING DATE: 2001-01-30	
31	PRIOR APPLICATION NUMBER: US 60/234,687	
32	PRIOR FILING DATE: 2000-09-21	
33	PRIOR APPLICATION NUMBER: US 09/608,408	
34	PRIOR FILING DATE: 2000-06-30	
35	PRIOR APPLICATION NUMBER: US 09/774,203	
36	PRIOR FILING DATE: 2001-01-29	
37	NUMBER OF SEQ ID NOS: 49117	
38	SOFTWARE: Annomax Sequence Listing Engine vers. 1.1	
39	SEQ ID NO 35629	
40	LENGTH: 42	
41	TYPE: PRT	
42	ORGANISM: Homo sapiens	
43	FEATURE:	
44	OTHER INFORMATION: MAP TO AC009492.1	
45	OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5	
46	OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4	
47	OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7	
48	OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3	
49	OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3	
50	OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2.4	
51	OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1	
52	OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6	
53	OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.3	
54	OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.98	
55	OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2	
56	US-09-864-761-35629	

Query Match 27.9%; Score 77; DB 9; Length 42;
Best Local Similarity 50.0%; Pred. No. 0.06;
Matches 13; Conservative 6; Mismatches 7; Indels

Qy 1 DNDDGGDNDGGGNNGGDGNATD 26
|: ||||:| |:|:|:|:|:|:
Db 17 DDGGDDDDDDNGDDGGDDDDVGD 42

Search completed: December 3, 2003, 16:24:58
Job time : 30 secs


```
QY 1 DNDGDD--NDGCGGNNGGDGNATDD 27
    ||||| :||| :||| :|||
Db 239 DNDDEDDKDDGGGDDGDDDEDD 267

RESULT 2
ID CAQ2 MOUSE STANDARD; PRT; 415 AA.
AC O09161; O88505;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calsequestrin, cardiac muscle isoform precursor (Calsequestrin 2).
GN CASQ2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RX MEDLINE=99014233; PubMed=9795116;
RA Park K.-W., Goo J.H., Chung H.-S., Kim H., Kim D.-H., Park W.-J.;
RT "Cloning of the genes encoding mouse cardiac and skeletal
RT calsequestrins: expression pattern during embryogenesis.";
RL Gene 217:25-30(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Sato Y., Ferguson D.G., Sako H., Dorn G.W. II, Kadambi V.J.,
RA Yatani A., Hoit B.D., Walsh R.A., Kranias E.G.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CALSEQUESTRIN IS A HIGH-CAPACITY, MODERATE AFFINITY,
CC CALCIUM-BINDING PROTEIN AND THUS ACTS AS AN INTERNAL CALCIUM STORE
CC IN MUSCLE. THE RELEASE OF CALCIUM BOUND TO CALSEQUESTRIN THROUGH
CC A CALCIUM RELEASE CHANNEL TRIGGERS MUSCLE CONTRACTION. BINDS 40
CC TO 50 MOLES OF CALCIUM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: THIS ISOFORM OF CALSEQUESTRIN OCCURS IN THE
CC SARCOPLASMIC RETICULUM'S TERMINAL CISTERNAE LUMINAL SPACES OF
CC CARDIAC AND SLOW SKELETAL MUSCLE CELLS.
CC -!- TISSUE SPECIFICITY: HEART.
CC -!- SIMILARITY: BELONGS TO THE CALSEQUESTRIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U91483; AAC82512.1; -.
CC EMBL; AF068244; AAC69472.1; -.
CC HSSP; P07221; 1A8Y.
CC MGD; MGI:1309469; Casq2.
CC InterPro; IPR001393; Calsequestrin.
CC Pfam; PF01216; Calsequestrin; 1.
CC PRINTS; PR00312; CALSEQUESTRIN.
CC PROSITE; PS00863; CALSEQUESTRIN 1; 1.
CC PROSITE; PS00864; CALSEQUESTRIN 2; 1.
CC SIGNAL; 1 19
KW SIGNAL; Glycoprotein; Calcium-binding; Signal.
FT SIGNAL 1 19
FT CHAIN 20 415
FT DOMAIN 373 415
FT CATH 335 335
FT CARBOHYD 1 2
FT CONFLICT 20 2
FT CONFLICT 21 2
FT CONFLICT 278 278
FT CONFLICT 415 AA; A3997ED046F8BD7E CRC64;
SQ SEQUENCE 415 AA; 48197 MW;

Query Match 26.88; Score 74; DB 1; Length 415;
Best Local Similarity 48.18; Pred. No. 0.27; Indels 0; Gaps 0;
Matches 13; Conservative 4; Mismatches 10;

QY 1 DNDGDDNDGCGGNNGGDGNATDD 27
    ||||| :||| :||| :|||
Db 383 DGDNDNDGDDDDDDDDDDNDSDEDD 409

RESULT 3
ID NNPI MOUSE STANDARD; PRT; 494 AA.
AC P56183; Q35712; Q9EREL; Q9J107; Q9JKU2;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NNP-1 protein (Novel nuclear protein 1) (Nops2).
GN NNPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RX TISSUE=Pituitary; PubMed=9192856;
RA Jansen E., Meulemans S.M.P., Orlans I.C.R., Van de Ven W.J.M.;
RT "The NNP-1 gene (D21S2056E), which encodes a novel nuclear protein,
RT maps in close proximity to the cystatin B gene within the EPM1 and
RT APECED critical region on 21q22.3.";
RL Genomics 42:336-341(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pituitary;
RA Jansen E.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=129/Ola; TISSUE=spleen;
RA Tsyba L., Skripkina I., Anoprienko O., Kvasha S., Slavov D.,
RA Tassone F., Ryndich A., Gardiner K.;
RT "Mouse homologues of human chromosome 21 genes.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS A CRITICAL ROLE IN THE GENERATION OF 28S RRNA (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=P56183-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P56183-2; Sequence=VSP_004335, VSP_004336;
CC Name=3;
CC IsoId=P56183-3; Sequence=VSP_004337;
CC -!- DEVELOPMENTAL STAGE: EMBRYONIC, EXPRESSION STARTING BETWEEN DAYS 1
CC AND 10.
CC -!- SIMILARITY: BELONGS TO THE NNP-1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U79773; AAC53286.1; -.
CC EMBL; U79774; AAB63261.1; -.
CC EMBL; AF294729; AAF76216.2; -.
CC EMBL; AF312394; AAG30293.1; -.
CC MGD; MGI:1203500; Nnpi.
CC Nuclear protein; Alternative splicing.
FT DOMAIN 247 305
FT VARSPLIC 245 258
FT ASDGDDGEASDGD -> GSEDDGDDSDDEA (in
FT isoform 2).
FT /FTId=VSP_004335.
```



```

FT VARSPLIC 259 309 Missing (in isoform 2).
FT /FTid=VSP_004336.
FT VARSPLIC 450 484 Missing (in isoform 3).
FT /FTid=VSP_004337.
FT CONFLICT 50 50 D -> C (IN REF. 3; AAG30293).
SQ SEQUENCE 494 AA; 54776 MW; 9866F1BD7AE3E9D CRC64;

Query Match 26.6%; Score 73.5; DB 1; Length 494;
Best Local Similarity 53.6%; Pred. No. 0.36;
Matches 15; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

QY 1 DNDGDDNDGGGE-NNGGDDNYATDD 27
|||:|||||:|||||:|||||:
Db 255 DGDGGEASDDDDGEASGGDDVADSDD 282

RESULT 4
YAB9 YEAST
ID YAB9 YEAST STANDARD; PRT; 1131 AA.
AC P31350;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 128.5 kDa helicase in ATSI-TPD3 intergenic region.
GN YAL019W OR YAL001 OR FUN30.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=93209532; PubMed=8458570;
RA Ouellette B.F.F., Clark M.W., Keng T., Storms R.K., Zhong W.W.,
RA Zeng B., Fortin N., Delaney S., Barton A.B., Kaback D.B., Bussey H.;
RT "Sequencing of chromosome I from Saccharomyces cerevisiae: analysis
RT of a 32 kb region between the LTEL and SPO7 genes.";
RL Genome 36; 32-42(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94193531; PubMed=8144453;
RA Barton A.B., Kaback D.B.;
RT "Molecular cloning of chromosome I DNA from Saccharomyces cerevisiae:
RT analysis of the genes in the FUN38-MAK16-SPO7 region.";
RL J. Bacteriol. 176:1872-1880(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=92221690; PubMed=1561836;
RA Clark M.W., Zhong W.W., Keng T., Storms R.K., Barton A.B.,
RA Kaback D.B., Bussey H.;
RT "Identification of a Saccharomyces cerevisiae homolog of the SNF2
RT transcriptional regulator in the DNA sequence of an 8.6 kb region in
RT the LTEL-CYS1 interval on the left arm of chromosome I.";
RL Yeast 8:133-145(1992).
CC -!- FUNCTION: MAY FUNCTION AS A NEGATIVE REGULATORY ELEMENT IN THE
CC PROCESS OF UV DAMAGE REPAIR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L05146; AAC04938.1;
CC PIR; S22266; S22266
CC SGD; S0000017; FUN30.
CC GO; GO:0007001; P:chromosome organization and biogenesis (sen. .; IGI.

```

```

DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR000330; SNF2 N.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00176; SNF2 N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
KW Hypothetical protein; Nuclear protein; DNA-binding; Helicase;
KW ATP-binding.
FT NP_BIND 597 604 ATP (POTENTIAL).
FT SITE 703 706 DECH BOX.
SQ SEQUENCE 1131 AA; 128506 MW; 7B61B9ABB3A42ED2 CRC64;

Query Match 26.4%; Score 73; DB 1; Length 1131;
Best Local Similarity 52.0%; Pred. No. 0.95;
Matches 13; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 3 DGDGDDNDGGGNGGDDNYATDD 27
|||:|||||:|||||:|||||:
Db 407 DDDDDNDGNNNNNNNTAGAD 431

RESULT 5
VG18 HSV11
ID VG18 HSV11 STANDARD; PRT; 318 AA.
AC Q00120;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Hypothetical gene 18 protein.
GN 18.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Ictalurid Herpes-like viruses.
OX NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Auburn 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M75136; AAA88121.1;
CC PIR; A36788; A36788.
KW Hypothetical protein.
FT DOMAIN 293 318 ASP-RICH (ACIDIC).
SQ SEQUENCE 318 AA; 36047 MW; A3DFB57E57CEE61 CRC64;

Query Match 25.4%; Score 70; DB 1; Length 318;
Best Local Similarity 54.2%; Pred. No. 0.57;
Matches 13; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

QY 1 DNDGDDNDGGG-GGNNNGGDDN 22
|||:|||||:|||||:|||||:
Db 294 DGDGDDGDDGDDGDDGDDGDDDD 317

RESULT 6
IE68 PRVKA
ID IE68 PRVKA STANDARD; PRT; 364 AA.
AC P24827;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)

```

```

DE GN Immediate-early protein RSP40.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037977; PubMed=2172457;
RA Zhang G., Leader D.P.;
RT "The structure of the pseudorabies virus genome at the end of the
RT inverted repeat sequences proximal to the junction with the short
RT unique region.";
RL J. Gen. Virol. 71:2433-2441(1990).
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
CC HSV-2 IE-68 (US1), EHV-1 65, EHV-4 (ORF4), PRV RSP40, AND VZV 63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00633; -; NOT ANNOTATED_CDS.
DR PIR; A36255; EDBESP.
DR InterPro; IPR003403; IE68.
DR Pfam; PF02479; IE68; 1.
KW Early protein.
FT DOMAIN 215 344 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 364 AA; 39613 MW; CE1329ECD7E1313 CRC64;

Query Match 25.4%; Score 70; DB 1; Length 364;
Best Local Similarity 44.4%; Pred. No. 0.65; Mismatches 9; Indels 0; Gaps 0;
Matches 12; Conservative 6;

Qy 1 DNDGDDNDGGGNGGNGGNGYATDD 27
| | | | | : | | | | : | | | | :
Db 310 DEDGDEDEDEGGDEGGDEGEDE 336

RESULT 7
YGG6 YEAST STANDARD; PRT; 657 AA.
ID YGG6 YEAST STANDARD; PRT; 657 AA.
AC P53165;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 72.9 kDa protein in RPB9-ALG2 intergenic region.
GN YGL066W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII.";
RL Yeast 13:1077-1090(1997).
RN [2]
RP SEQUENCE OF 115-657 FROM N.A.
RX STRAIN=S288C;
RA Feuerhahn M., de Montigny J., Potier S., Souciet J.-L.;
RT "The characterization of two new clusters of duplicated genes
RT suggests a 'lego' organization of the yeast Saccharomyces cerevisiae
RT chromosomes.";
RL Yeast 13:861-869(1997).
CC -!- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02247; AAA41949.1; -.
DR EMBL; M11898; AAA41958.1; -.

```

DR PIR: A03296; PIRT3.
 KW Repeat; Parotid gland; Multigene family; Signal.
 FT SIGNAL 1 13
 FT CHAIN 14 206
 FT DOMAIN 80 189
 FT REPEAT 80 97
 FT REPEAT 98 115
 FT REPEAT 116 133
 FT REPEAT 134 152
 FT REPEAT 153 170
 FT REPEAT 171 189
 SQ SEQUENCE 206 AA; 21403 MW; D037582CED05CE6B CRC64;
 Query Match 25.0%; Score 69; DB 1; Length 206;
 Best Local Similarity 52.4%; Pred. No. 0.47;
 Matches 11; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 OY 1 DNDGDDNDGCGENNGDGD 21
 DB 55 ENGDDSDGDDGDDGSDGDN 75
 RESULT 9
 ID TYVI HUMAN STANDARD; PRT; 414 AA.
 AC P25490; Q14935;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
 DE transcription factor) (NF-E1).
 GN YY1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=9205716; PubMed=1655281;
 RA Shi Y., Seto E., Chang L.-S., Shenk T.;
 RA "Transcriptional repression by YY1, a human GLI-Kruppel-related
 RT protein, and relief of repression by adenovirus E1A protein.";
 RL Cell 67:377-388(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=ForeSkin;
 RC MEDLINE=92052179; PubMed=1946405;
 RA Park K., Atchison M.;
 RA "Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta),
 RT that binds to the immunoglobulin kappa 3' enhancer and the
 RL immunoglobulin heavy-chain mu E1 site.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9804-9808(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Whitson R.H., Huang T., Dang J., Itakura K.;
 RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J.J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=98152931; PubMed=9493912;
 RA McNeil S., Guo B., Stein J.L., Lian J.B., Bushmeyer S., Seto E.,
 RA Atchison M.L., Penman S., van Wijnen A.J., Stein G.S.;
 RT "Targeting of the YY1 transcription factor to the nucleolus and the
 RT nuclear matrix in situ: the C-terminus is a principal determinant for
 RT nuclear trafficking.";
 RL J. Cell. Biochem. 68:500-510(1998).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 304-414.
 RX MEDLINE=97098436; PubMed=8942976;
 RA Houbaviv H.B., Ushva A., Shenk T., Burley S.K.;
 RT "Cocrysal structure of YY1 bound to the adeno-associated virus P5
 RT initiator.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13577-13582(1996).
 RN [7]
 RP STRUCTURE BY NMR OF 353-379.
 RX MEDLINE=98308000; PubMed=9642075;
 RA Viles J.H., Patel S.U., Mitchell J.B.O., Moody C.M., Justice D.E.,
 RA Uppenbrink J., Doyle P.M., Harris C.J., Sadler P.J., Thornton J.M.;
 RT "Design, synthesis and structure of a zinc finger with an artificial
 RT beta-turn.";
 RL J. Mol. Biol. 279:973-986(1998).
 CC -1- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
 CC POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
 CC VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
 CC START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
 CC DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
 CC REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. FOR
 CC EXAMPLE IT ACTS AS A REPRESSOR IN ABSENCE OF ADENOVIRUS E1A
 CC PROTEIN BUT AS AN ACTIVATOR IN ITS PRESENCE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M77698; AAA59467.1; -;
 DR EMBL; M76541; AAA59926.1; -;
 DR EMBL; Z14077; CAA78455.1; -;
 DR EMBL; BC037308; AAH37308.1; -;
 DR PIR; A40350; A40350.
 DR PDB; 1UBD; 23-DEC-96.
 DR PDB; 1ZNN; 01-APR-98.
 DR TRANSFAC; T00915; -;
 DR Genew; HGNC:12856; YY1.
 DR MIM; 600013; -;
 DR GO; GO:0003713; F:transcription co-activator activity; TAS.
 DR GO; GO:0003714; F:transcription co-repressor activity; TAS.
 DR GO; GO:0003700; F:transcription factor activity; TAS.
 DR GO; GO:0008270; F:zinc ion binding activity; TAS.
 DR GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .; TAS.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf_C2H2; 1.
 DR ProDom; PD000003; Znf_C2H2; 1.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.

KW Transcription regulation; Repressor; Activator; Nuclear protein;
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat; 3D-structure.
 FT DOMAIN 43 53 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 54 69 GLY-RICH.
 FT DOMAIN 70 80 POLY-HIS.
 FT DOMAIN 159 170 GLY/SER-RICH.
 FT ZN_FING 296 320 C2H2-TYPE.
 FT ZN_FING 325 347 C2H2-TYPE.
 FT ZN_FING 353 377 C2H2-TYPE.
 FT ZN_FING 383 407 C2H2-TYPE.
 FT DOMAIN 257 341 INVOLVED IN NUCLEAR MATRIX ASSOCIATION.
 FT DOMAIN 333 371 INVOLVED IN REPRESSION OF ACTIVATED
 TRANSCRIPTION.
 FT DOMAIN 371 397 INVOLVED IN MASKING TRANSACTIVATION
 DOMAIN.
 FT CONFLICT 65 65 H -> R (IN REF. 2).
 FT CONFLICT 196 196 G -> R (IN REF. 1).
 FT STRAND 296 297
 FT TURN 301 302
 FT STRAND 306 307
 FT HELIX 310 317
 FT STRAND 318 320
 FT STRAND 325 326
 FT TURN 328 330
 FT STRAND 333 334
 FT STRAND 337 343
 FT HELIX 344 346
 FT TURN 347 347
 FT STRAND 353 354
 FT TURN 358 359
 FT STRAND 363 364
 FT HELIX 367 378
 FT TURN 388 389
 FT TURN 397 400
 FT HELIX 401 407
 SQ SEQUENCE 414 AA; 44712 MW; 058C05A0AD2D04E6 CRC64;

Query Match 25.0%; Score 69; DB 1; Length 414;
 Best Local Similarity 45.8%; Pred. No. 0.96;
 Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
 QY 1 DNDGDDNDGGGNGGQDNYA 24
 Db 45 EEDDDDDGGGGGGGGHGA 68

RESULT 10
 YUY2_YEAST
 ID YUY2_YEAST STANDARD; PRT; 207 AA.
 AC P47087;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 23.2 kDa protein in SPCL-ILV3 intergenic region.
 GN YJ012C OR J1440 OR YJ083.25.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / FY1679;
 RA De Haan M., Smits P.H.M., Grivell L.A.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X87611; CAA60934.1; -;
 DR EMBL; Z49512; CAA89536.1; -;
 DR PIR; S55200; S55200.
 DR SGD; S0003773; YJR012C.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 17 37 POTENTIAL.
 SQ SEQUENCE 207 AA; 23190 MW; FDC11E6ED5AFA90A CRC64;
 Query Match 24.8%; Score 68.5; DB 1; Length 207;
 Best Local Similarity 50.0%; Pred. No. 0.54;
 Matches 13; Conservative 4; Mismatches 6; Indels 3; Gaps 1;
 QY 2 NDDGDDNDGGGNGGQDNYATDD 27
 Db 174 NDSNDNDNDNDNNKGGND---DD 196
 RESULT 11
 KEX2_CANAL
 ID KEX2_CANAL STANDARD; PRT; 924 AA.
 AC O13359;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Kexin precursor (EC 3.4.21.61) (KEX2 protease).
 GN KEX2.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WO-1;
 RX MEDLINE=98030572; PubMed=9360967;
 RA Newport G.R., Agabian N.;
 RT "KEX2 influences Candida albicans proteinase secretion and hyphal
 RT formation.";
 RL J. Biol. Chem. 272:28954-28961(1997).
 CC -!- CATALYTIC ACTIVITY: Cleavage of Lys-Arg-|-Xaa and Arg-Arg-|-Xaa
 CC bonds to process Yeast alpha-factor pheromone and killer toxin
 CC precursors.
 CC -!- COFACTOR: CALCIUM DEPENDENT (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. LATE GOLGI
 CC COMPARTMENT (BY SIMILARITY).
 CC -!- PTM: O-GLYCOSYLATED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF023372; AAB80929.1; -;
 DR HSP; P00782; 2SBT.
 DR MEROPS; S08.070; -;
 DR InterPro; IPR002884; P_domain.
 DR InterPro; IPR002009; Peptidase_S8.
 DR Pfam; PF01483; P_protein; PARTIAL.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR ProDom; PD000717; P_domain; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Calcium; Transmembrane;
 KW Zymogen; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 ?
 FT CHAIN ? 924 KEXIN.
 FT DOMAIN ? 774 LUMENAL (POTENTIAL).

[illegible]

```

DR PROSITE; PS00863; CALSQUESTRIN 1; 1.
DR PROSITE; PS00864; CALSQUESTRIN 2; 1.
KW Muscle; Glycoprotein; Calcium-binding; Signal.
FT SIGNAL 1 19
FT CHAIN 20 409
FT DOMAIN 373 409
FT CARBOHYD 335 335
FT CONFLICT 45 45
SQ SEQUENCE 409 AA; 47356 MW; 068D469EA7457DAE CRC64;

Query Match 24.3%; Score 67; DB 1; Length 409;
Best Local Similarity 44.4%; Pred. No. 1.6;
Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 DNDGDDNDGGGNGGDDNYATDD 27
Db 382 DDDDDDDDDGNSDEDD 408

RESULT 14
CH16 POPTR
ID _CH16 POPTR STANDARD; PRT; 340 AA.
AC P16579;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Acidic endochitinase WIN6 precursor (EC 3.2.1.14).
GN WIN6.
OS Populus trichocarpa (Western balsam poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3694;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H11-11; TISSUE=Leaf;
RX MEDLINE=9435563; PubMed=8075397;
RA Clarke H.R., Davis J.M., Wilbert S.M., Bradshaw H.D. Jr., Gordon M.P.;
RT "Wound-induced and developmental activation of a poplar tree
RT chitinase gene promoter in transgenic tobacco.";
RL Plant Mol. Biol. 25:799-815(1994).
RN [2]
RP SEQUENCE OF 127-340 FROM N.A.
RX MEDLINE=90046703; PubMed=2813366;
RA Parsons T.J., Bradshaw H.D. Jr., Gordon M.P.;
RT "Systemic accumulation of specific mRNAs in response to wounding in
RT poplar trees.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7895-7899(1989).
CC -1- FUNCTION: THIS PROTEIN FUNCTIONS AS A DEFENSE AGAINST CHITIN
CC CONTAINING FUNGAL PATHOGENS.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -1- INDUCTION: By wounding.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS IA WHICH IS SIMILAR TO
CC CLASS IB IN THE CATALYTIC PART BUT WHICH INCLUDE A N-TERMINAL
CC CHITIN-BINDING LECTIN DOMAIN (BELONGS TO FAMILY 19 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U01661; AAA57278.1; -.
CC EMBL; U01660; AAA57277.1; -.
CC EMBL; M25336; AAA96701.1; -.
CC HSSP; P02877; IHEV
CC InterPro; IPR001002; Chitin binding.1.
DR InterPro; IPR000726; Glyco_Hydro_19.
DR Pfam; PF00187; Chitin_bind_1; 1.

```

```

DR Pfam; PF00182; Glyco_hydro_19; 1.
DR PRINTS; PR00451; CHITINBINDNG.
DR ProDom; PD000609; Chitin_binding_1; 1.
DR ProDom; PD354900; Glyco_hydro_19; 1.
DR SMART; SM00270; ChcBD1; 1.
DR PROSITE; PS00026; CHITIN_BINDING; 1.
DR PROSITE; PS00773; CHITINASE_19_1; 1.
DR PROSITE; PS00774; CHITINASE_19_2; 1.
KW Hydrolase; Glycosidase; Chitin degradation; Chitin-binding;
KW Multigene family; Signal.
FT SIGNAL 1 22
FT CHAIN 23 340
FT DOMAIN 23 62
FT DOMAIN 63 85
FT DOMAIN 86 340
FT DISULFID 25 40
FT DISULFID 34 46
FT DISULFID 39 53
FT DISULFID 57 61
FT DISULFID 110 172
FT DISULFID 183 191
FT DISULFID 290 323
SQ SEQUENCE 340 AA; 36409 MW; C4E96473BEAA55C5 CRC64;

Query Match 24.1%; Score 66.5; DB 1; Length 340;
Best Local Similarity 56.0%; Pred. No. 1.5;
Matches 14; Conservative 4; Mismatches 4; Indels 3; Gaps 2;

QY 5 GDNDNDG--GGENNGGDDGN-YATD 26
Db 67 GDGDDGDCGDDGDDGDDGYLSD 91

RESULT 15
CAQ2 RAT
ID _CAQ2 RAT STANDARD; PRT; 413 AA.
AC P51868; O09177;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calsequestrin, cardiac muscle isoform precursor (Calsequestrin 2).
GN CASQ2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Heart muscle;
RA Aquilla T.T., Rovner A.S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 50-413 FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Rodriguez M.M., Chen C., Smith B., Mochly-Rosen D.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CALSQUESTRIN IS A HIGH-CAPACITY, MODERATE AFFINITY,
CC CALCIUM-BINDING PROTEIN AND THUS ACTS AS AN INTERNAL CALCIUM STORE
CC IN MUSCLE. THE RELEASE OF CALCIUM BOUND TO CALSQUESTRIN THROUGH
CC A CALCIUM RELEASE CHANNEL TRIGGERS MUSCLE CONTRACTION. BINDS 40
CC TO 50 MOLES OF CALCIUM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: THIS ISOFORM OF CALSQUESTRIN OCCURS IN THE
CC SARCOPLASMIC RETICULUM'S TERMINAL CISTERNAE LUMINAL SPACES OF
CC CARDIAC AND SLOW SKELETAL MUSCLE CELLS.
CC -1- SIMILARITY: BELONGS TO THE CALSQUESTRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 16:16:14 ; Search time 35 Seconds
(without alignments)
390.765 Million cell updates/sec

Title: US-09-991-003B-8

Perfect score: 276

Sequence: 1 DNDGDNDGGGNGGDDG.....MMMVVVVVVVVVIIIMLLMT 53

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:.*
1: sp_archea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	36.1	1360	5 Q94649	Q94649 plasmodium
2	92	33.3	1249	5 Q812R2	Q812R2 plasmodium
3	85.5	31.0	1864	5 Q81CP5	Q81CP5 plasmodium
4	85	30.8	899	5 Q81BN8	Q81BN8 plasmodium
5	85	30.8	2961	5 Q81452	Q81452 plasmodium
6	84.5	30.6	894	5 Q81511	Q81511 plasmodium
7	82	29.7	734	10 Q9LN20	Q9LN20 arabidopsis
8	81.5	29.5	1176	5 Q8SSV4	Q8SSV4 dictyosteli
9	81	29.3	1308	5 Q81R00	Q81R00 plasmodium
10	80	29.0	90	4 Q8BY99	Q8BY99 homo sapien
11	80	29.0	2276	2 Q93TY6	Q93TY6 staphylococ
12	78.5	28.4	1559	5 Q81406	Q81406 plasmodium
13	78	28.3	447	12 Q9VTJ3	Q9VTJ3 ateline her
14	78	28.3	966	5 Q81LK5	Q81LK5 plasmodium
15	77.5	28.1	4095	5 Q81CN0	Q81CN0 plasmodium
16	77	27.9	393	5 Q26592	Q26592 schistosoma

17	77	27.9	764	3 Q12500	Q12500 saccharomyc
18	77	27.9	1946	5 Q97291	Q97291 plasmodium
19	76.5	27.7	607	10 Q9FG45	Q9FG45 arabidopsis
20	76.5	27.7	1105	5 Q81M35	Q81M35 plasmodium
21	76.5	27.7	1431	5 Q81LH5	Q81LH5 plasmodium
22	76	27.5	222	5 Q96829	Q96829 plasmodium
23	76	27.5	461	12 Q9J861	Q9J861 spodoptera
24	75.5	27.4	669	5 Q8XS88	Q8XS88 drosophila
25	75.5	27.4	669	5 Q9VY66	Q9VY66 drosophila
26	75.5	27.4	1641	5 Q81K20	Q81K20 plasmodium
27	75.5	27.4	5439	5 Q81586	Q81586 plasmodium
28	75	27.2	382	10 Q9LK24	Q9LK24 arabidopsis
29	75	27.2	945	5 Q814Y8	Q814Y8 plasmodium
30	75	27.2	2162	5 Q81BH2	Q81BH2 plasmodium
31	75	27.2	2754	5 Q81451	Q81451 plasmodium
32	74.5	27.0	1371	5 Q81CK4	Q81CK4 plasmodium
33	74	26.8	189	5 Q9NH20	Q9NH20 drosophila
34	74	26.8	420	5 Q81JE1	Q81JE1 plasmodium
35	74	26.8	1125	5 Q81BU0	Q81BU0 plasmodium
36	74	26.8	1173	5 Q8MS50	Q8MS50 drosophila
37	74	26.8	5779	5 Q81BS0	Q81BS0 plasmodium
38	73.5	26.6	299	10 Q8HJD9	Q8HJD9 oryza sativ
39	73.5	26.6	2820	5 Q81M32	Q81M32 plasmodium
40	73	26.4	297	5 Q81HT0	Q81HT0 plasmodium
41	73	26.4	366	5 Q815P9	Q815P9 plasmodium
42	73	26.4	411	11 Q8CHJ4	Q8CHJ4 rattus norv
43	73	26.4	541	10 Q9MA97	Q9MA97 arabidopsis
44	73	26.4	682	5 Q81I98	Q81I98 plasmodium
45	73	26.4	780	5 Q81I22	Q81I22 plasmodium

ALIGNMENTS

RESULT 1

Q94649 ID Q94649 PRELIMINARY; PRT; 1360 AA.
AC Q94649;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE AARP2 protein.
GN AARP2.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=97391121; PubMed=9247928;
RA Barale J.C., Attal-Bonnefoy G., Brahimi K., Pereira L., Langsley G.;
RT "AARP2 is a novel evolutionary conserved protein whose repeats
RT identify a new antigen family in Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 87:169-181(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=97378065; PubMed=9234746;
RA Barale J.C., Candelle D., Attal-Bonnefoy G., Dehoux P., Bonnefoy S.,
RA Ridley R., Pereira da Silva L., Langsley G.;
RT "Plasmodium falciparum AARP1, a giant protein containing repeated
RT motifs rich in asparagine and aspartate residues, is associated with
RT the infected erythrocyte membrane.";
RL Infect. Immun. 65:3003-3010(1997).
DR EMBL; Y08924; CAA70129.1; -.
DR InterPro; IPR007034; DUF663.
DR Pfam; PF04950; DUF663; 1.
SQ SEQUENCE 1360 AA; 158109 MW; 400FF6AC9C74AE3A CRC64;

Query Match 36.1%; Score 99.5; DB 5; Length 1360;
Best Local Similarity 40.4%; Pred. No. 0.001;
Matches 23; Conservative 8; Mismatches 15; Indels 11; Gaps 2;

Q9BY99	PRELIMINARY;	PRT;	90 AA.
Q9BY99;			
AC			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	Hypochemical protein KIAA1661 (Fragment).		
DE	KIAA1661.		
GN	homo sapiens (Human).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Hirosawa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;		
RT	"Identification of the transcribed sequences on human chromosome 22 by		
RT	the expressed sequence tags-mapping.";		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DDJ databases.		
DR	EMBL; AB051448; BAB33331.1; -.		

Q81406	PRELIMINARY;	PRT: 1559 AA.
ID	Q81406	
AC	Q81406;	
DT	01-MAR-2003 (TrEMBLrel. 23, Created)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Phosphatidylinositol 4-kinase, putative (EC 2.7.1.67).	
GN	PF0485W.	
OS	Plasmodium falciparum (isolate 3D7).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium	
OX	NCBI_TaxID=36329;	
OX	[1]	
RN	SEQUENCE FROM N.A.	
RP		
RA	Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pai	
RA	Hall N., Bowman S., Churcher C., Quail M., Barrall B.;	

RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=22255708; PubMed=12368867;	
RA	Hall N., Pain A., Barriman M., Churcher C., Harris B., Harris D.,	
RA	Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,	
RA	Buckee C.O., Burrows C., Chervach I., Chillingworth C.,	
RA	Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,	
RA	Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,	
RA	Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,	
RA	Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,	
RA	Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,	
RA	Knighte A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,	
RA	Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,	
RA	Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,	
RA	Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,	
RA	Seeger K., Sharp S., Smith R., Squares R., Squares M., Stevens K.,	
RA	Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,	
RA	Sulston J.E., Craig A., Newbold C., Barrrell B.G;	
RT	"Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";	
RL	Nature 419:527-531(2002).	
DR	EMBL; AL929352; CA051463.1;	
KW	Kinase; Transferase.	
SQ	SEQUENCE 1559 AA; 182663 MW; 9087140E1E176E96 CRC64;	
Query Match 28.4%; Score 78.5; DB 5; Length 1559;		
Best Local Similarity 35.4%; Pred. No. 0.45;		
Matches 17; Conservative 7; Mismatches 11; Indels 13; Gaps 1;		
QY	1 DNDGDDNDGGGNNGGDGN-----YATDDMMVMIMM 35	
Db	131 DDDGDDGDDGDDGDDGDDSDSDSDSDIKNNHKNYNDYHKNMIGL 178	
RESULT 13		
Q9YTJ3	PRELIMINARY;	
ID	Q9YTJ3	PRT; 447 AA.
AC	Q9YTJ3;	
DT	01-MAY-1999 (TrEMBLrel. 10, Created)	
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Oxf 73.	
OS	Ateeline herpesvirus 3.	
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;	
OC	Gammapherpesvirinae; Rhadinovirus.	
OX	NCBI_TaxID=85618;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=73;	
RX	MEDLINE=20091363; PubMed=106233770;	
RA	Albrecht J.C.;	
RT	"Primary structure of the Herpesvirus Ateles genome.";	
RL	J. Virol. 74:1033-1037(2000).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=73;	
RA	Albrecht J.-C., Fleckenstein B.;	
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF083424; AAC95598.1;	
SQ	SEQUENCE 447 AA; 46504 MW; EA242FE09A6A5EB9 CRC64;	
Query Match 28.3%; Score 78; DB 12; Length 447;		
Best Local Similarity 65.0%; Pred. No. 0.14;		
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;		
QY	1 DNDGDDNDGGGNNGGDG 20	
Db	176 DGEDGDDGDDGGGDDGG 195	
RESULT 14		
Q8ILKS	PRELIMINARY;	
ID	Q8ILKS	PRT; 966 AA.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 16:16:44 ; Search time 21 Seconds
(without alignments)
242.711 Million cell updates/sec

Title: US-09-991-003B-8
Perfect score: 276
Sequence: 1 DNDGDDNDGGGNGGDDG.....MMMVVVVIMVVMVIMLLMT 53

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76.*

1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	36.1	1360	2 T18403	asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum)
2	79	28.6	334	2 A29561	prostatic spermine
3	78	28.3	447	2 T42987	hypothetical prote
4	77	27.9	764	2 S64951	hypothetical prote
5	73	26.4	1131	2 S22286	FUN30 protein - ye
6	72	26.1	91	2 F69069	hypothetical prote
7	72	26.1	297	2 T30613	hypothetical prote
8	70.5	25.5	1948	2 S00485	gene 11-1 protein
9	70	25.4	318	2 A36788	hypothetical prote
10	70	25.4	364	1 EDBESP	hypothetical prote
11	70	25.4	498	2 D71616	immediate-early pr
12	70	25.4	657	2 S64073	ribosome releasing
13	70	25.4	3394	2 T18501	hypothetical prote
14	69.5	25.2	156	2 S52246	hypothetical prote
15	69	25.0	170	2 A48013	transposable retro
16	69	25.0	206	1 PIR73	proline-rich prote
17	69	25.0	367	2 JC4831	acidic proline-ric
18	69	25.0	414	2 A40350	adsorption protein
19	69	25.0	3844	2 T18402	transcription repr
20	68.5	24.8	207	2 S55200	asparagine/asparta
21	68.5	24.8	485	2 T49237	hypothetical prote
22	68	24.6	94	2 B84539	hypothetical prote
23	67.5	24.5	426	2 T05676	hypothetical prote
24	67.5	24.5	782	2 S59387	probable membrane
25	67.5	24.5	1553	2 T18502	hypothetical prote
26	67	24.3	210	2 T40719	hypothetical prote
27	67	24.3	295	2 B48013	hypothetical prote
28	67	24.3	409	1 JQ1396	caldesquestrin prec
29	67	24.3	638	2 T13691	hypothetical prote

30 67 24.3 1360 2 F96596 hypothetical prote
31 67 24.3 1755 2 T51532 hypothetical prote
32 66.5 24.1 340 2 S48030 probable chitinase
33 66.5 24.1 4550 2 T18440 hypothetical prote
34 66 23.9 147 2 JQ1320 high mobility grou
35 66 23.9 316 2 G96576 hypothetical prote
36 66 23.9 513 2 B96524 hypothetical prote
37 66 23.9 644 2 S63056 probable membrane
38 66 23.9 1089 2 S48241 NMD2 protein - yea
39 66 23.9 1094 2 S46021 probable regulator
40 66 23.9 2269 2 T18472 hypothetical prote
41 65.5 23.7 168 2 C86388 hypothetical prote
42 65.5 23.7 872 2 T18861 probable peptida
43 65 23.6 117 2 T45695 hypothetical prote
44 65 23.6 674 2 S66937 probable membrane
45 65 23.6 998 2 I41078 hemolysin - Escher

ALIGNMENTS

RESULT 1

T18403
asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18403
R:BARALE, J.C.; ATTALBONNEFOY, G.; BRAHIMI, K.; DASILVA, L.P.; LANGSLEY, G.
Mol. Biochem. Parasitol. 87, 169-181, 1997
A:Title: Plasmodium-falciparum asparagine and aspartate rich protein-2 is an evolution
A:Reference number: Z08834; MUID:97391121; PMID:9247928
A:Accession: T18403
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1360 <BAR>
A:Cross-references: EMBL:Y08924; NID:el008106; PID:el154303; PIDN:CAA70129.1
C:Genetics:
A:Gene: aarp2

Query Match 36.1%; Score 99.5; DB 2; Length 1360;
Best Local Similarity 40.4%; Pred. No. 0.0023;
Matches 23; Conservative 8; Mismatches 15; Indels 11; Gaps 2;

QY 4 DGDNDGCGGNGGCGD-----DNYATDD--MMMVIMMVVVIMVIM 49
DB 401 DGDNDGCGGNGGCGD-----DNYATDD--MMMVIMMVVVIMVIM 457

RESULT 2

A29561
prostatic spermine-binding protein precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Dec-1998 #sequence_revision 15-Dec-1988 #text_change 05-Nov-1999
C:Accession: A29561
R:Chang, C.; Saltzman, A.G.; Hilpakka, R.A.; Huang, I.Y.; Liao, S.
J. Biol. Chem. 262, 2826-2831, 1987
A:Title: Prostatic spermine-binding protein. Cloning and nucleotide sequence of cDNA,
A:Reference number: A29561; MUID:87137539; PMID:3818623
A:Accession: A29561
A:Molecule type: mRNA
A:Residues: 1-334 <CHA>
A:Cross-references: GB:J02675; NID:g206855; PIDN:AAA42113.1; PID:g206856

Query Match 28.6%; Score 79; DB 2; Length 334;
Best Local Similarity 55.2%; Pred. No. 0.12;
Matches 16; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 1 DNDGDD--NDDCGGNGGCGDNYATDD 27
DB 294 DNDGDDKDDCGGCGDGDGDDDDDD 322

RESULT 3


```
A;Residues: 1-297 <SEN>
A;Cross-references: EMBL:U60315; PIDN:AAC55139.1
C;Genetics:
A;Note: MC011L

Query Match      26.1%; Score 72; DB 2; Length 297;
Best Local Similarity 48.1%; Pred. No. 0.63;
Matches 13; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 DNDGDDNDGCGGNNGGDGNATDD 27
Db 183 DGDGCGGCGGNGGCGGCGGCGGDDDD 209

RESULT 8
S00485
gene 11-1 protein precursor - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jun-2000
C;Accession: S00485
R;Scherf, A.; Hilbich, C.; Sieg, K.; Mattei, D.; Mercereau-Pujalon, O.; Mueller-Hill, B
EMBO J. 7, 1129-1137, 1988
A;Title: The 11-1 gene of Plasmodium falciparum codes for distinct fast evolving repeats
A;Reference number: S00485; MUID:88296416; PMID:284111
A;Accession: S00485
A;Molecule type: DNA
A;Residues: 1-1315;1316-1485;1486-1657;1658-1729;1730-1948 <SCH>
A;Cross-references: EMBL:X07453
C;Comment: This protein is associated with the membrane of red blood cells at the schiz
C;Genetics:
A;Gene: 11-1
A;Introns: 71/3
C;Keywords: tandem repeat
F;1-71/Domain: signal sequence #status predicted <SIG>
F;72-1948/Product: gene 11-1 protein (fragments) #status predicted <MAT>

Query Match      25.5%; Score 70.5; DB 2; Length 1948;
Best Local Similarity 53.6%; Pred. No. 5.7;
Matches 15; Conservative 2; Mismatches 8; Indels 3; Gaps 1;

Qy 2 NDDGDDN---DDGCGGNNGGDGNATD 26
Db 269 NTEGDKGVEKGGGEGGCGHGSAGD 296

RESULT 9
A36788
hypothetical protein ORF18 - ictaluriid herpesvirus 1 (strain auburn 1)
C;Species: ictaluriid herpesvirus 1
A;Note: host ictalurius punctatus (channel catfish)
C;Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 08-Oct-1999
C;Accession: A36788
R;Davison, A.J.
submitted to GenBank, January 1992
A;Description: Channel catfish virus: a new type of herpesvirus.
A;Reference number: A36804
A;Accession: A36788
A;Molecule type: DNA
A;Residues: 1-318 <DAV>
A;Cross-references: GB:W751336; NID:G331209; PIDN:AAA88121.1; PID:G331228
R;Davison, A.J.
Virology 186, 9-14, 1992
A;Title: Channel catfish virus: a new type of herpesvirus.
A;Reference number: A39447; MUID:92087490; PMID:1727613
A;Contents: annotation
A;Note: neither protein nor nucleic acid sequence is given
C;Genetics:
A;Gene: 18

Query Match      25.4%; Score 70; DB 2; Length 318;
Best Local Similarity 54.2%; Pred. No. 1.1;
Matches 13; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

A;Residues: 1-657 <RIE>
```

```
Qy 1 DNDGDDNDGCGGNNGGDGN 22
Db 294 DGDGCGGDDGDDGDDGDDGDDDD 317

RESULT 10
EDBESF
immediate-early protein RSP40 - suid herpesvirus 1 (strain Ka)
C;Species: suid herpesvirus 1
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 08-Apr-1994
C;Accession: A36255
R;Zhang, G.; Leader, D.P.
J. Gen. Virol. 71, 2433-2441, 1990
A;Title: The structure of the pseudorabies virus genome at the end of the inverted ref
A;Reference number: A36255; MUID:91037977; PMID:2172457
A;Accession: A36255
A;Molecule type: DNA
A;Residues: 1-364 <ZHA>
C;Superfamily: herpesvirus immediate-early protein IE68
C;Keywords: early protein

Query Match      25.4%; Score 70; DB 1; Length 364;
Best Local Similarity 44.4%; Pred. No. 1.3;
Matches 12; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 DNDGDDNDGCGGNNGGDGNATDD 27
Db 310 DEDGDEDEDEGDEGDEGDEGDEDE 336

RESULT 11
D71616
ribosome releasing factor (OO, TP) PFB0390w [imported] - malaria parasite (Plasmodium
N;Alternate names: ribosome recycling factor
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 15-Sep-2000
C;Accession: D71616
R;Gardner, M.J.; Tetzelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V
; Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: D71616
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-498 <GAR>
A;Cross-references: GB:AE001391; GB:AE001362; NID:G3845168; PIDN:AACT1867.1; PID:G3845
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0390w

Query Match      25.4%; Score 70; DB 2; Length 498;
Best Local Similarity 51.9%; Pred. No. 1.7;
Matches 14; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 1 DNDGDDNDGCGGNNGGDGNATDD 27
Db 130 DRUDDDDNYDDDDDDNYDDDDDD 156

RESULT 12
S64073
hypothetical protein YGL066w - Yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein G3406
C;Species: Saccharomyces cerevisiae
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
C;Accession: S64073; S64070
R;Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64071
A;Accession: S64073
A;Molecule type: DNA
A;Residues: 1-657 <RIE>
```

A;Cross-references: EMBL:Z72588; NID:g1322573; PIDN:CAA96770.1; PID:e243288; PID:g132257
A;Experimental source: strain S288C
R;Feuermann, M.; Potier, S.; Souciet, J.L.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64044
A;Accession: S64070
A;Molecule type: DNA
A;Residues: 115-657 <FEU>
A;Cross-references: EMBL:Z72588; MIPS:YGL066w
A;Experimental source: strain S288C
C;Genetics:
A;Cross-references: SGD:S0003034
A;Map position: 7L

Query Match 25.4%; Score 70; DB 2; Length 657;
Best Local Similarity 40.7%; Pred. No. 2.3;
Matches 11; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 1 DNDGDDDDGGGNGGDDNVATDD 27
|||||:||||: : : : :
DB 147 DNDNEDDDDDDDDDDDNSNGANYKKND 173

RESULT 13
T18501
hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18501
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18501
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3394 <LAW>
A;Cross-references: EMBL:Z98551; NID:el331903; PID:el331922; PIDN:CAB1140.1
C;Genetics:
A;Map position: 3
A;Note: C0760c

Query Match 25.4%; Score 70; DB 2; Length 3394;
Best Local Similarity 44.4%; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 DNDGDDDDGGGNGGDDNVATDD 27
|||||:||||: : : : :
DB 3010 DNDNEDDDDDDDDDDDNNDDNNDDDD 3036

RESULT 14
S52246
transposable retroelement homolog - perennial teosinte
N;Alternate names: ZLRS7 protein, long repetitive sequence protein ZLRS 7
C;Species: Zea diploperennis (perennial teosinte)
C;Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 07-Dec-1999
C;Accession: JC4641; S52246
R;Monfort, A.; Vicient, C.M.; Raz, R.; Puigdomenech, P.; Martinez-Izquierdo, J.A.
DNA Res. 2, 255-261, 1995
A;Title: Molecular analysis of a putative transposable retroelement from the Zea genus
A;Reference number: JC4641; MUID:97021439; PMID:8867799
A;Accession: JC4641
A;Molecule type: DNA
A;Residues: 1-156 <MON>
A;Cross-references: EMBL:X82087; NID:g609287; PIDN:CAA57619.1; PID:g609288
F;123-153/Region: DNA binding #status predicted

Query Match 25.2%; Score 69.5; DB 2; Length 156;
Best Local Similarity 54.2%; Pred. No. 0.65;
Matches 13; Conservative 4; Mismatches 4; Indels 3; Gaps 1;

QY 1 DNDGDDDDGGGNGG---DGD 21
|:|||||:| : : : :|

Db 95 DDDGDDDDGGGDDGGSGGGRSSGD 118

RESULT 15
A48013
proline-rich proteoglycan 1 precursor, parotid - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 20-Aug-1999
C;Accession: A48013
R;Castle, A.M.; Castle, J.D.
J. Biol. Chem. 268, 20490-20496, 1993
A;Title: Novel secretory proline-rich proteoglycans from rat parotid. Cloning and char.
A;Reference number: A48013; MUID:93388626; PMID:8376404
A;Accession: A48013
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-170 <CAS>
A;Cross-references: GB:L17317; NID:g310197; PIDN:AAA03073.1; PID:g310198
C;Superfamily: proline-rich protein
C;Keywords: extracellular protein; glycoprotein; tandem repeat

Query Match 25.0%; Score 69; DB 2; Length 170;
Best Local Similarity 52.4%; Pred. No. 0.8;
Matches 11; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 DNDGDDDDGGGNGGDDGD 21
:|||||:| : : : :|
DB 55 ENGDDDDDDDDGGGDDGN 75

Search completed: December 3, 2003, 16:20:05
Job time : 22 secs